



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

TP 2: Análisis de Redes Sociales

Octubre 27, 2022

Métodos Numéricos

Grupo 18

| Integrante | LU | Correo electrónico |
|-------------------|--------|----------------------|
| Vekselman, Natán | 338/21 | natanvek11@gmail.com |
| Arienti, Federico | 316/21 | fa.arianti@gmail.com |
| Manuel Lakowsky | | |
| Brian Kovo | | |



Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (+54 +11) 4576-3300

<http://www.exactas.uba.ar>

RESUMEN

La descomposición de matrices en autovectores y autovalores aparece en una variedad de aplicaciones donde importa caracterizar el comportamiento de un sistema: en el reconocimiento de imágenes, en el análisis de estabilidad de cuerpos rotantes, en el análisis de riesgo de mercado y en el análisis de redes —por nombrar algunas—. Desde un punto de vista geométrico, se puede considerar a los autovectores como los ‘ejes’ de una transformación lineal, en tanto representan una dirección invariante a la transformación, y a los autovalores como los factores por los que esas direcciones se comprimen, estiran o invierten.

En este trabajo proponemos una implementación en C++ de un método para el cálculo de autovalores y autovectores en matrices cuadradas. Para algunas matrices particulares, como pueden ser las matrices simétricas definidas positivas, este método nos permitirá obtener todos sus autovalores y autovectores asociados. El mismo se conoce como *el método de la potencia con deflación*.

A su vez, presentaremos dos aplicaciones concretas de los autovalores y autovectores en el análisis de redes: la medición de centralidad de autovector y corte mínimo en la red del ‘Club de Karate’ [5], y la estimación de una *ego-network* [3] de Facebook, por medio de la construcción de una matriz de similaridad.

Palabras clave: *método de la potencia, deflación de Hotelling, centralidad de autovector, conectividad algebraica, análisis de componentes principales.*

CONTENIDOS

| | |
|---|----|
| 1. Método de la potencia con deflación | 2 |
| 1.1. Introducción teórica | 2 |
| 1.2. Implementación | 3 |
| 1.3. Evaluación cuantitativa | 5 |
| 2. Análisis: Club de Karate | 7 |
| 2.1. Contexto | 7 |
| 2.2. Centralidad de Autovector | 8 |
| 2.3. Autovectores de la matriz laplaciana | 9 |
| 3. Análisis: Red ‘Ego’ | 13 |
| 3.1. Contexto | 13 |
| 3.2. Matriz de similaridad | 13 |
| 3.3. Comparación con la red original | 13 |
| 3.4. Optimización | 13 |
| 3.5. PCA | 13 |
| 4. Conclusiones | 14 |
| 5. Apéndice | 15 |
| Referencias | 16 |

1. MÉTODO DE LA POTENCIA CON DEFLACIÓN

1.1. Introducción teórica. El método de la potencia con deflación permite aproximar un subconjunto de los autovalores y autovectores asociados a una matriz. Si la misma satisface que todos sus autovalores son no nulos y diferentes en módulo, entonces permite aproximar el conjunto entero.

MÉTODO DE LA POTENCIA: el método de la potencia, *Power method* o *Power iteration*, es una técnica iterativa para aproximar el autovector asociado al autovalor en módulo máximo de una matriz cuadrada que satisfaga esta característica —es decir, tenga un autovalor dominante no nulo—, a partir de la aplicación de sucesivos productos matriciales, descriptos por la siguiente relación de recurrencia:

$$b_0 \text{ es un vector aleatorio : } ||b_0|| = 1$$

$$(1) \quad b_{k+1} = \frac{\mathbf{A}b_k}{||\mathbf{A}b_k||}$$

donde $|| \cdot ||$ es una norma vectorial.

Se puede demostrar [1] que, bajo las condiciones descriptas, si b_0 no es ortogonal al autovector asociado al autovalor dominante en módulo de \mathbf{A} , b_k convergerá a éste. Lo que es más, se podrá aproximar el autovalor dominante por medio del coeficiente de Rayleigh:

$$(2) \quad \lambda_{max} = \frac{b_k^t \mathbf{A} b_k}{b_k^t b_k}$$

MÉTODO DE LA DEFLACIÓN: el método de la deflación, por su parte, corresponde a la transformación de la matriz inicial \mathbf{A} por una matriz \mathbf{B} con autovalores equivalentes, salvo por el autovalor dominante que será anulado. Existen distintos métodos de deflación, entre ellos la deflación de Hotelling y la deflación de Wielandt [1].

En este trabajo utilizaremos la deflación de Hotelling por su sencillez, a costas de un mayor error numérico [1]. El mismo consiste en aplicar el método de la potencia para sucesivas matrices que satisfagan la siguiente relación:

$$\mathbf{B}_0 = \mathbf{A}$$

$$(3) \quad \mathbf{B}_{k+1} = \mathbf{B}_k - \lambda v v^t$$

donde λ corresponde al autovalor en módulo máximo estimado por el método de la potencia y v es su autovector asociado.

Se puede demostrar que \mathbf{B}_{k+1} contiene los mismos autovalores que \mathbf{B}_k , salvo el máximo que quedará anulado.

1.2. Implementación. Procedemos a detallar una posible implementación para ambos métodos, restringiéndonos al caso de autovalores reales. Definimos:

$$\text{deflacion} : \text{matriz}_{n \times n} \mathbf{A} \times \text{nat } k \times \text{nat } q \times \text{real } t \longrightarrow \text{vector}_k \times \text{matriz}_{n \times k}$$

$$\text{potencia} : \text{matriz}_{n \times n} \mathbf{A} \times \text{nat } q \times \text{real } t \longrightarrow \text{real} \times \text{vector}_n$$

donde n es un natural, \mathbf{A} tiene al menos k autovalores reales dominantes en módulo, $0 < k \leq n$, q es un número par¹ que representa el máximo de iteraciones a realizar y $0 \leq t$ representa la tolerancia mínima a partir de la que se considera la convergencia de una solución.

```

1 proc deflacion(in A: matriz<n, n>, in k: Nat, in q: Nat, in t: Real) {
2
3     eigvals := vector<q>
4     eigvecs := matriz<n, k>
5
6     i := 0
7     while i < k {
8
9         l, V := potencia(A, q, t)
10        eigvals[i] := l
11        eigvecs.columna[i] := V
12
13        A := A - l * (V * V.t)
14        i := i + 1
15    }
16
17    return eigvals, eigvecs
18 }
```

ALGORITMO 1. Pseudocódigo para el método de la deflación.

El algoritmo (1.) retornará un vector con los primeros k autovalores en módulo máximos de \mathbf{A} , ordenados descendientemente, y una matriz cuyas columnas corresponden, respectivamente, a los autovectores normalizados asociados a estos autovalores.

Es interesante notar que la k -ésima matriz sobre la que se aplicará el método de la potencia — \mathbf{B}_k — tendrá, por definición, un autovalor cero con multiplicidad algebraica mayor o igual

¹Esta restricción no es necesaria, pero permite realizar una optimización que se explica en la próxima página.

a k . En el caso en que la matriz inicial sea simétrica, \mathbf{B}_k será simétrica² y, en consecuencia, diagonalizable. Se puede demostrar que la única matriz diagonalizable con multiplicidad algebraica $\mu_a(0) = n$ es la matriz nula, por lo que el método de la deflación de Hotelling tenderá hacia ella. Esto nos permite inferir que el error numérico será proporcional a k^3 . Es decir, los autovalores más chicos de la matriz serán más susceptibles a errores.

Por su parte, el algoritmo (2.) retornará el autovalor de \mathbf{A} máximo en módulo y su autovector asociado:

```

1 proc potencia(in A: matriz<n, n>, in q: Nat, in t: Real) {
2
3     q := q / 2
4     B := A * A
5
6     v := aleatorio(n)      // un vector aleatorio no nulo
7     v := v / norma(v, 2)  // ||v||_2 = 1
8
9     i := 0
10    while i < q {
11
12        y := B * v
13        y := y / norma(y, 2)
14        if (norma(v - y, 2) < t) { // tolerancia para la convergencia
15            break
16        }
17        v := y
18        i := i + 1
19    }
20
21    l := (v.t * A * v) / (v.t * v) // coeficiente de rayleigh
22
23    return l, v
24 }
```

ALGORITMO 2. Pseudocódigo para el método de la potencia.

Una primera observación importante es que el algoritmo (2.) no es capaz de distinguir si la selección inicial del vector v es ortogonal al autovector asociado al autovalor en módulo dominante, por lo que una mala selección puede resultar en que el algoritmo falle. Para reducir la probabilidad de ocurrencia, se propone —heurísticamente— que se elija cada coordenada del vector de manera pseudo-aleatoria sobre un rango amplio. Por ejemplo, la máxima representación de enteros con signo en 32 bits.

²Por suma de simétricas. Notar que vv^t siempre resulta en una matriz de este tipo.

³En tanto existirá una correlación. Sin embargo, es esperable que otros factores entren en juego: la variabilidad de los autovalores, el número de condición de la matriz, la selección del vector aleatorio inicial, o la cantidad de iteraciones q a realizar, por ejemplo.

Además, proponemos un modelo diferente al que define la relación de recurrencia de la ecuación (1.). Su convergencia, en realidad, depende del signo del autovalor en módulo dominante. Si este es negativo, la secuencia es acotada. Se puede demostrar⁴ que la subsecuencia definida por $\{b_k : k \text{ es par}\}$ siempre convergerá. El algoritmo (2.) aprovecha estas observaciones.

Desde un punto de vista temporal, también permite reducir el costo de ejecución, ya que se iterará a lo sumo $q/2$ veces. Estimamos que la complejidad de *potencia* estará en $\Theta(n^3 + \frac{q}{2}n^2)$.

1.3. Evaluación cuantitativa. Procedemos a evaluar nuestra implementación del método de la potencia con deflación en C++ acorde a los algoritmos propuestos.

ERROR RELATIVO. medimos el error⁵ $\|\mathbf{A}\mathbf{V} - \mathbf{V}\mathbf{\Lambda}\|_1$ y $\|\mathbf{A}\mathbf{V} - \mathbf{V}\mathbf{\Lambda}\|_\infty$ para 30 instancias de matriz $\mathbf{A} \in \mathbb{R}^{100 \times 100}$ generadas aleatoriamente, donde \mathbf{V} y $\mathbf{\Lambda}$ representan — respectivamente — las matrices aproximadas de autovectores y autovalores de \mathbf{A} , tal que $\mathbf{A}\mathbf{V}_i \approx \mathbf{\Lambda}_{ii}\mathbf{V}_i \quad \forall i : 1 \dots n$.

METODOLOGÍA: se calculó $\mathbf{\Lambda}, \mathbf{V} = \text{deflacion}(\mathbf{A}, 100, 50000, 1e-10)$ y se midió el error relativo para cada una de las matrices.

Cada caso se generó a través de uno de los siguientes tres procedimientos⁶:

- 1) *Matrices Diagonales:* Se generaron mil matrices diagonales \mathbf{D} con mil autovalores en el rango $[-1e4, 0) \cup (0, 1e4]$. Los mismos se generaron con el rng *PCG64* de numpy para evitar distribuciones particulares que pudieran influir en la variabilidad de los autovalores.
- 2) *Matrices Diagonalizables:* Se generaron mil matrices diagonalizables $\mathbf{A} := \mathbf{Q}\mathbf{D}\mathbf{Q}^t$ donde cada matriz \mathbf{D} se generó a partir de la metodología (1.) y $\mathbf{Q} := \mathbf{I} - 2uu^t$ se generó a partir de un vector aleatorio u —con el algoritmo `random.rand()` de numpy— tal que $\|u\|_2 = 1$.
- 3) *Matrices Simétricas Definidas Positivas:* Se generaron mil matrices simétricas definidas positivas de enteros. Para cada caso se generó una matriz aleatoria \mathbf{B} con el algoritmo `random.randint()` de numpy en el rango $[-1e4, 0) \cup (0, 1e4]$ y se procedió a definir $\mathbf{A} := \mathbf{B}\mathbf{B}^t$.

OBSERVACIONES: consideramos que la varianza de los autovalores, el número de condición de las matrices, y su tamaño, son variables que afectan de manera significativa en el error del procedimiento. El proceso mencionado para la generación de matrices aleatorias fue pensado

⁴Esto se desprende de la demostración de convergencia de la secuencia original, donde la convergencia de \mathbf{B}_k depende de $(\lambda/|\lambda|)^k = \text{signo}(\lambda)^k$, con λ el autovalor en módulo dominante.

⁵El script asociado se puede encontrar en `./experimentos/error_potencia.py`.

⁶Se utilizó un valor semilla para facilitar la reproductibilidad.

sobre generadores de números aleatorios para tratar de minimizar cualquier tendencia que pueda provenir de la utilización de distribuciones particulares. Se controló el tamaño n , la cantidad de iteraciones q y la tolerancia t . Un estudio más exhaustivo debería también evaluar el comportamiento del algoritmo en función de estos parámetros.

RESULTADOS. [\[experimento\]](#)

2. ANÁLISIS: CLUB DE KARATE

2.1. **Contexto.** La red del *Club de Karate* fue parte de una investigación antropológica [5], realizada por Wayne W. Zachary, que estudió las relaciones ‘políticas’ entre los miembros de un club universitario. La misma se realizó durante el desarrollo de un conflicto que terminó por dividir al grupo.

La red buscó modelar el flujo de información entre sus integrantes por medio de la tripla $(\mathbf{V}, \mathbf{E}, \mathbf{C})$, donde \mathbf{V} denota el conjunto de individuos, \mathbf{E} refiere a un grafo no direccionado cuyos ejes representan los vínculos entre los miembros del club, y \mathbf{C} define la fuerza de estas relaciones —lo que se podría pensar como ponderaciones sobre los ejes de \mathbf{E} —.

La investigación tuvo como enfoque central demostrar la capacidad del modelo para predecir la división del grupo por medio del algoritmo de *labeling* de ‘flujo máximo - corte mínimo’ de Ford y Fulkerson [2].

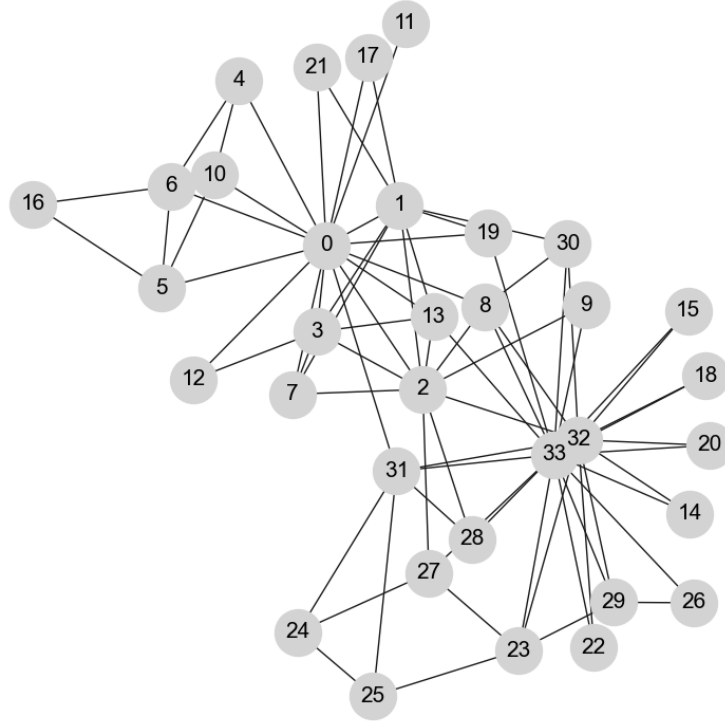


FIGURA 1. La red del *Club de Karate*. Cada nodo representa un individuo, cada eje la existencia de un vínculo por fuera del club.

En este análisis⁷ utilizaremos sólo la representación matricial de \mathbf{E} para evaluar la importancia de los distintos miembros en la red, y la matriz laplaciana asociada para evaluar el uso de autovectores como predictores de la división del grupo.

⁷El script asociado se puede encontrar en `./experimentos/club_de_karate.py`, los archivos con los resultados en `./experimentos/resultados/club-karate/`.

2.2. Centralidad de Autovector. La centralidad de autovector es una medida que se utiliza en el análisis de redes para medir la ‘importancia’ de los nodos que componen una red, relativa a la importancia de sus conexiones. Dada una matriz de conectividad \mathbf{W} , se define:

$$(4) \quad \lambda x = \mathbf{W}x$$

donde λ es el autovalor en módulo máximo de \mathbf{W} .

Intuitivamente, se puede pensar que la importancia de cada nodo es proporcional a la suma de las importancias de sus vecinos. Se puede demostrar [4] que, dado las características de esta matriz, el autovector asociado siempre será positivo en coordenadas.

En tanto la red de *Club de Karate*, podemos ver que la aplicación del metodo de la potencia sobre la matriz de conectividad asociada al grafo \mathbf{E} resulta en el siguiente autovector asociado al autovalor en módulo máximo:

| | | | |
|----|----------|----|----------|
| 0 | 0.355491 | 17 | 0.092400 |
| 1 | 0.265960 | 18 | 0.101403 |
| 2 | 0.317193 | 19 | 0.147913 |
| 3 | 0.211180 | 20 | 0.101403 |
| 4 | 0.075969 | 21 | 0.092400 |
| 5 | 0.079483 | 22 | 0.101403 |
| 6 | 0.079483 | 23 | 0.150119 |
| 7 | 0.170960 | 24 | 0.057052 |
| 8 | 0.227404 | 25 | 0.059206 |
| 9 | 0.102674 | 26 | 0.075579 |
| 10 | 0.075969 | 27 | 0.133477 |
| 11 | 0.052856 | 28 | 0.131078 |
| 12 | 0.084255 | 29 | 0.134961 |
| 13 | 0.226473 | 30 | 0.174758 |
| 14 | 0.101403 | 31 | 0.191034 |
| 15 | 0.101403 | 32 | 0.308644 |
| 16 | 0.023636 | 33 | 0.373363 |

FIGURA 2. Centralidad de autovector para la red del *Club de Karate*. La columna izquierda denota el nodo, la derecha su ‘importancia’.

Vemos que el nodo ‘0’ y el nodo ‘33’ son los más centrales. Esto no es casualidad, la red del *Club de Karate* está armada para tener a las dos figuras principales del conflicto en cada extremo —el instructor de karate y el presidente del club, respectivamente— para satisfacer la especificación del algoritmo de labeling que utiliza.

2.3. Autovectores de la matriz laplaciana. La matriz laplaciana $\mathbf{L} = \mathbf{D} - \mathbf{W}$ —donde \mathbf{D} es una matriz diagonal con elementos $d_{ii} = \sum_j w_{ij}$ y \mathbf{W} es una matriz de conectividad— sirve para medir distintas propiedades en una red.

En particular, el mínimo autovalor en módulo no nulo —llamado de conectividad algebraica, o valor de Fiedler— permite establecer un criterio sobre el que particionar la red en dos. El autovector asociado a este autovalor designará la pertenencia de un nodo a una u otra partición acorde a su signo.

Procedemos a analizar qué autovector de la matriz laplaciana asociada a \mathbf{E} permite predecir mejor la división que ocurrió en el *Club de Karate*. Para ello, calculamos los autovectores de la matriz laplaciana con el método de la potencia con deflación y medimos el valor absoluto de la correlación entre cada autovector y un vector que indica la división verdadera que ocurrió en el grupo.

| | | | |
|-----------|----------|----------|------------------|
| 18.136696 | 0.266429 | 3.013963 | 0.055720 |
| 17.055171 | 0.066721 | 2.749157 | 0.940899 |
| 13.306122 | 0.457220 | 2.487092 | 0.684434 |
| 10.921068 | 0.339844 | 2.0 | 0.0 |
| 9.777241 | 0.496086 | 2.0 | 0.0 |
| 6.996197 | 0.060813 | 2.0 | 0.0 |
| 6.515545 | 0.461652 | 2.0 | 0.0 |
| 6.331592 | 0.082665 | 2.0 | 0.0 |
| 5.618034 | 0.0 | 1.955050 | 0.065144 |
| 5.378595 | 1.427676 | 1.826055 | 0.401069 |
| 4.580793 | 0.465710 | 1.761899 | 0.158209 |
| 4.480008 | 0.262230 | 1.599283 | 0.325133 |
| 4.275877 | 0.062841 | 1.259404 | 0.002377 |
| 3.472187 | 0.001872 | 1.125011 | 1.943500 |
| 3.381966 | 0.0 | 0.909248 | 1.550553 |
| 3.376154 | 0.379940 | 0.468525 | 4.750632 |
| 3.242067 | 0.505417 | 0.0 | nan ⁸ |

FIGURA 3. Correlación entre los autovectores asociados a los autovalores de la red del *Club de Karate* y un vector que indica la división verdadera que ocurrió en el grupo. La columna izquierda denota el autovalor, la derecha la correlación.

Vemos que el valor de conectividad algebraica —el autovalor mínimo en módulo no nulo— tiene asociado el mejor autovector para realizar el corte. En la figura (4.) se puede observar que sólo dos nodos obtuvieron una clasificación incorrecta: el ‘2’ y el ‘8’, lo que presenta un desempeño casi igual al logrado en la investigación original⁹.

⁸El método de la potencia no está definido si el autovalor máximo en módulo es nulo. Este resultado da cuenta de ello. Como nos interesan solo los autovalores diferentes a ‘0’, optamos por descartarlo.

⁹En este sentido, solo el nodo ‘2’ recibió una clasificación incorrecta. Sobre el ‘8’, Zachary considera [5] que hubo un interés ‘egoísta’, por parte del individuo asociado a ese nodo, que lo hizo actuar de manera distinta a su ‘afiliación’ política (es decir, sus relaciones).

| | | | | | | | |
|----|-----------|---|---|----|-----------|---|---|
| 0 | -0.112137 | 0 | 0 | 17 | -0.100181 | 0 | 0 |
| 1 | -0.041288 | 0 | 0 | 18 | 0.162751 | 1 | 1 |
| 2 | 0.023219 | 1 | 0 | 19 | -0.013637 | 0 | 0 |
| 3 | -0.055500 | 0 | 0 | 20 | 0.162751 | 1 | 1 |
| 4 | -0.284605 | 0 | 0 | 21 | -0.100181 | 0 | 0 |
| 5 | -0.323727 | 0 | 0 | 22 | 0.162751 | 1 | 1 |
| 6 | -0.323727 | 0 | 0 | 23 | 0.155695 | 1 | 1 |
| 7 | -0.052586 | 0 | 0 | 24 | 0.153026 | 1 | 1 |
| 8 | 0.051601 | 1 | 0 | 25 | 0.160963 | 1 | 1 |
| 9 | 0.092801 | 1 | 1 | 26 | 0.187110 | 1 | 1 |
| 10 | -0.284605 | 0 | 0 | 27 | 0.127664 | 1 | 1 |
| 11 | -0.210993 | 0 | 0 | 28 | 0.095152 | 1 | 1 |
| 12 | -0.109461 | 0 | 0 | 29 | 0.167650 | 1 | 1 |
| 13 | -0.014742 | 0 | 0 | 30 | 0.073500 | 1 | 1 |
| 14 | 0.162751 | 1 | 1 | 31 | 0.098753 | 1 | 1 |
| 15 | 0.162751 | 1 | 1 | 32 | 0.130345 | 1 | 1 |
| 16 | -0.422765 | 0 | 0 | 33 | 0.118903 | 1 | 1 |

FIGURA 4. Autovector de Conectividad algebraica, la primer columna representa los nodos de la red, la segunda el autovector, la tercera la clasificación de cada nodo acorde a su signo y la cuarta la división verdadera que ocurrió.

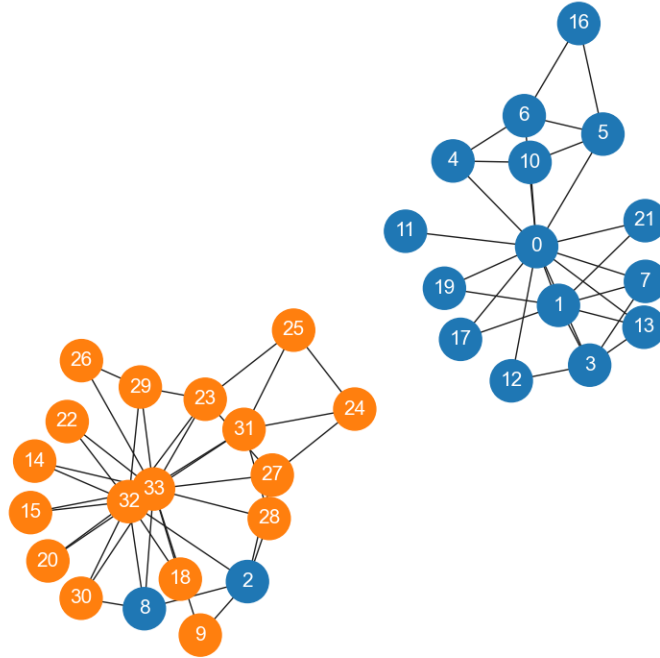
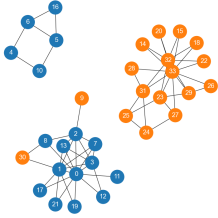
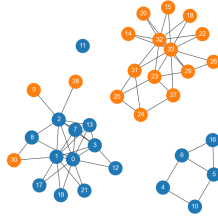


FIGURA 5. La división estimada de la red del *Club de Karate*. Los colores representan la división verdadera de la red. Azul: '0', Naranja: '1'.

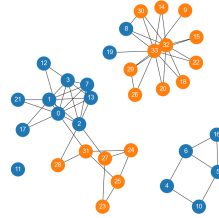
Procedemos a detallar —a modo de comparación visual— los cortes realizados por los demás autovectores. Notamos que el proceso de corte sólo elimina los ejes existentes entre pares cuya clasificación difiere. Por ello, la cantidad de componentes conexos que puede generar un corte no es necesariamente dos.



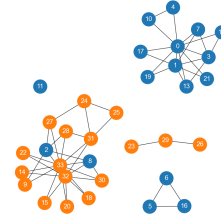
$av \approx 0.91$



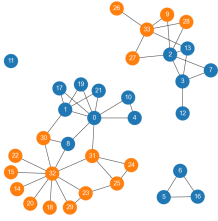
$av \approx 1.13$



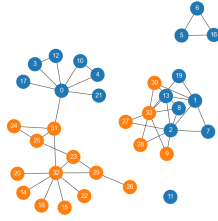
$av \approx 1.26$



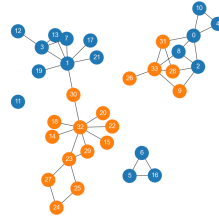
$av \approx 1.60$



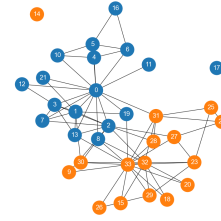
$av \approx 1.76$



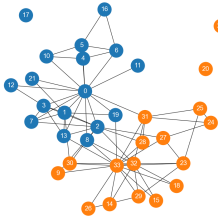
$av \approx 1.83$



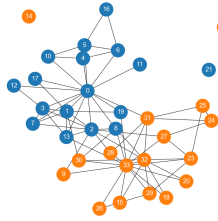
$av \approx 1.96$



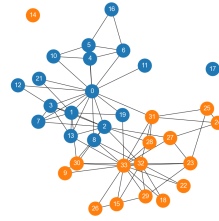
$av \approx 2.00$



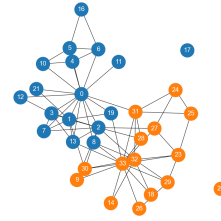
$av \approx 2.00$



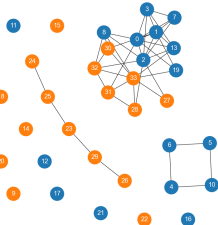
$av \approx 2.00$



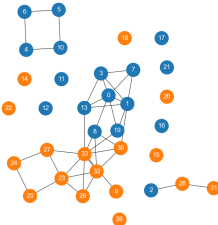
$av \approx 2.00$



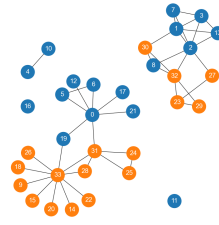
$av \approx 2.00$



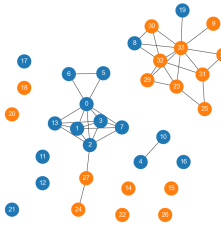
$av \approx 2.49$



$av \approx 2.75$



$av \approx 3.01$



$av \approx 3.24$

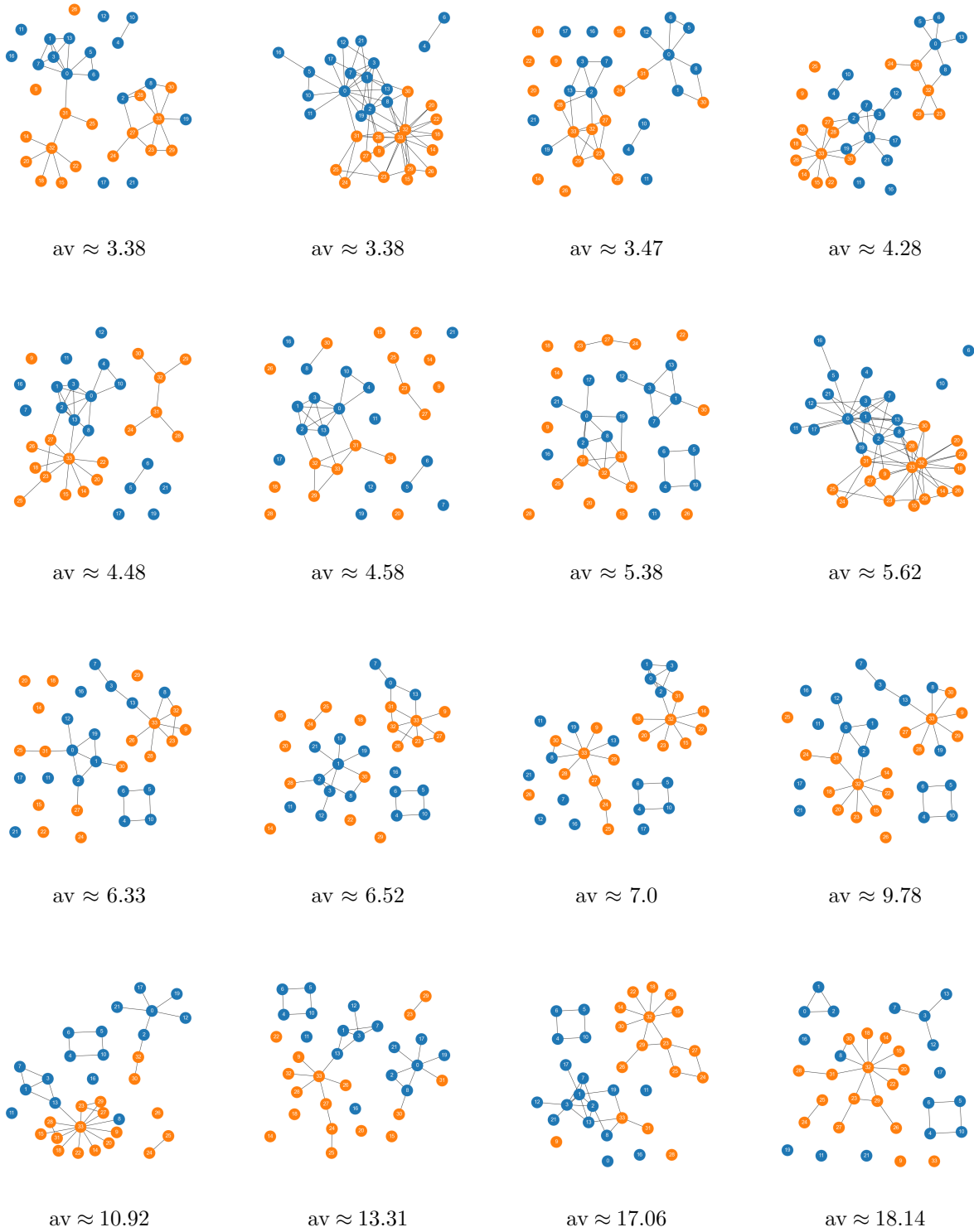


FIGURA 5. Todos los cortes posibles —salvo el de Fiedler y el nulo—, correspondientes a los autovectores asociados a la matriz laplaciana de la red del *Club de Karate*. Cada uno se designa por medio de su autovalor.

3. ANÁLISIS: RED ‘EGO’

3.1. Contexto. Una *ego-network* [3] es una red compuesta por las amistades que existen entre los amigos de un individuo, el ‘ego’. Estas redes son centrales para aplicaciones como Facebook, Google+ o Twitter.

En particular, dada una red ‘ego’, resulta de interés poder identificar los círculos sociales —conjuntos, disjuntos y anidados— a los que pertenece un usuario. Leskovec [3] propone un método de aprendizaje no supervisado para lograr inferirlos, que se nutre de la siguiente información: un grafo \mathbf{E} (la red) —donde se espera que exista una correlación fuerte entre un círculo y la densidad de conexiones entre los nodos que lo componen— y un conjunto de atributos \mathbf{C} , para cada nodo —donde se espera una correlación entre un círculo y la similitud de los atributos de los nodos que lo componen—.

En este análisis estimaremos \mathbf{E} , una red ‘ego’ proveniente de Facebook, por medio de la construcción de una matriz de similitud que utilice los atributos definidos en \mathbf{C} . También, buscaremos reducir su dimensionalidad por medio del análisis de componentes principales.

3.2. Matriz de similitud.

3.3. Comparación con la red original.

3.4. Optimización.

3.5. PCA.

4. CONCLUSIONES

5. APÉNDICE

REFERENCIAS

- [1] Richard L Burden and J Douglas Faires. *Numerical Analysis*. 2000.
- [2] L Ford and D Fulkerson. *Flows in networks*. Princeton University Press, 1962.
- [3] Jure Leskovec and Julian Mcauley. Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25, 2012.
- [4] Mark EJ Newman. The mathematics of networks. *The new palgrave encyclopedia of economics*, 2(2008):1–12, 2008.
- [5] Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.